基于改进SlowFast模型的设施黄瓜农事行为识别方法

何　峰[1,2]，吴华瑞[1,2,3,4]，史扬明[1,2]，朱华吉[1,2,3,4*]

（1. 江苏大学 计算机科学与通信工程学院，江苏镇江 212013，中国；2. 国家农业信息化工程技术研究中心，北京 100097，中国；3. 北京市农林科学院信息技术研究中心，北京 100097，中国；4. 农业农村部数字乡村技术重点实验室，北京 100097，中国）

摘　要：［**目的/意义**］农事行为活动识别对设施蔬菜生产精准化调控有着重要意义，在一定程度上可以通过查看农事操作的时间、操作过程是否合理来减少因农事行为不当导致产量下降。为了解决农事行为识别方法中由于黄瓜叶片和设施遮挡导致识别准确率不高的问题，提出一种名为SlowFast-SMC-ECA（SlowFast-Spatio-Temporal Excitation、Channel Excitation、Motion Excitation-Efficient Channel Attention）的农事活动行为识别算法。［**方法**］该算法主要基于SlowFast模型，通过改进Fast Pathway和Slow Pathway中的网络结构来提高对于农事活动中手部动作特征和关键特征的提取能力。在Fast Pathway中，引入多路径激励残差网络的概念，通过在信道之间插入卷积操作来增强它们在时域上的相互关联性，从而更好地捕捉快速运动信息的细微时间变化。在Slow Pathway中，将传统的Residual Block替换为ECA-Res结构，以提高对通道信息的捕获能力。这两项改进有效地加强了通道之间的联系，提升了特征之间的语义信息传递，进而显著提升了农事行为识别的准确率。此外，为了解决数据集中类别不均衡的问题，设计了平衡损失函数（Smoothing Loss），通过引入正则化系数，平衡损失函数可以有效地处理数据集中的类别不均衡情况，提高模型在各个类别上的表现。［**结果和讨论**］改进的SlowFast-SMC-ECA模型在农事行为识别中表现出良好的性能，各类行为的平均识别精度达到80.47%，相较于原始的SlowFast模型有约3.5%的提升。［**结论**］本研究在农事行为识别中展现出良好的性能。这对农业生产的智能化管理和决策具有重要意义。

关键词：农事活动行为；SlowFast模型；多路径激励残差网络；ECA-Res；平衡损失函数

中图分类号：TP391.4　　**文献标志码：**A　　**文章编号：**SA202402001

引用格式：何峰，吴华瑞，史扬明，朱华吉．基于改进SlowFast模型的设施黄瓜农事行为识别方法[J]. 智慧农业(中英文), 2024, 6(3): 118-127. DOI：10.12133/j.smartag.SA202402001

HE Feng, WU Huarui, SHI Yangming, ZHU Huaji. Recognition Method of Facility Cucumber Farming Behaviours Based on Improved SlowFast Model[J]. Smart Agriculture, 2024, 6(3): 118-127. DOI：10.12133/j. smartag. SA202402001 (in Chinese with English abstract)

0　引　言

黄瓜在中国各地普遍栽培，因含有丰富的人体所需的各种营养物质，对人体健康非常有益而深受消费者的喜欢[1,2]。在黄瓜栽培过程中会存在大量的农事活动行为，如浇水、吊蔓、剪枝等。这些农事操作得合理与否直接影响黄瓜的产量和品质，进而影响整个生产的产出效益。同时，农事操作的时间、操作过程、投入农资量、投入精准度等基础数据也是实现黄瓜生产精准化调控管理的依据，因此如何快速准确地记录农事操作行为就显得尤为重要。

传统的农事活动行为记录主要依靠人工进行。记录过程存在时间延迟、准确度不高、信息遗漏等问题。这些会给黄瓜的生产管理造成一定的影响。随着图像识别与计算机视觉技术的飞速发展，基于机器视觉技术，通过对农事活动行为视频的自动提取和识别实现农事活动记录成为一种可行的技术方案。

行为识别方法可以大致分为两类：一类是基于传统方法，需要手工提取和设计特征以进行识别；另一类则借助深度学习技术，通过神经网络自动学习数据中的特征，从而对一些简单的行为（如挥

收稿日期：2024-02-01

基金项目：中央引导地方科技发展资金项目（2023ZY1-CGZY-01）；财政部和农业农村部：国家现代农业产业技术体系资助（CARS-23-D07）

作者简介：何　峰，研究方向为计算机视觉，E-mail：1363263324@qq.com

*通信作者：朱华吉，博士，研究员，研究方向为农业信息化，E-mail：zhuhj@nercita.org.cn

手、聊天）进行识别[3]。手工特征提取方法主要是通过人工方法提取视频中的运动信息，然后使用分类器如支持向量积（Support Vector Machine, SVM）[4]、K临近算法[5]、贝叶斯分类器[6,7]等，对动作进行检测分类[8]。它们充分利用了运动物体的外观特征。这些特征不仅简单易懂，而且具有出色的鲁棒性。这种方法已经成为基于视频识别行为的优选，并且在多个领域得到了广泛的应用。手工特征的可行性和广泛性使其成为一个强大的工具，用于捕捉和分析视频中的运动、形状、颜色、纹理等关键信息，从而实现对行为的准确识别和分析。此外，一些学者认为视频图像携带着前后帧的运动信息，通过提取这些信息，可以计算出光流，进而获取图像中物体运动的光流数据，从而用于描述运动状态。如，Wang等[9]采用了一种密集轨迹法（Dense Trajectories, DT）的方法，通过在视频帧中密集提取轨迹点，并捕捉这些轨迹点随时间的变化，用于行为识别和动作分析。之后，Wang和Schmid[10]在DT的算法上进行了改进，提出了改进的密集轨迹法（Improved Dense Trajectories, IDT），通过更精细的轨迹采样和增强的特征提取技巧，提高了在视频中捕获动作信息的效率和准确性，使其在行为识别和动作分析中更具竞争力。

近年来，深度学习领域取得了迅猛发展，为行为识别研究提供了崭新的视角和方法。传统的手工特征提取方法通常伴随着内存需求较高的问题，并受到特征单一性的限制，从而在扩展性方面存在一定的挑战。这些深度学习方法不仅能够高效处理大规模数据，还能够自动从数据中学习丰富的特征表示，因此在视频行为识别等领域表现出巨大的潜力。主流的基于深度学习的视频理解算法包括双流卷积神经网络（Two-Stream Convolutional Neural Networks, Two-Stream CNN）、人体骨架识别、三维卷积神经网络（3D CNN），以及视觉Transformer。这些网络结构在捕捉视频中的行为特征和动作信息方面发挥着重要的作用。2014年，Simonyan和Zisserman[11]提出了一种创新的方法，即双流卷积神经网络。这个网络采用了两个分支：一个分支专门用于提取时间流特征；另一个分支则专注于提取空间流特征。在网络的后端，它将这两个流的特征融合在一起，以实现更加全面和高效的信息提取和表示。这一方法为视频行为识别等任务带来了重要的突破，使得模型能够更好地理解时间和空间信息，从而提高了识别性能。在此基础上进行改进的网络有TSN（Temporal Segment Networks）[12]网络和I3D（Inflated 3D ConvNet）[13]网络。3D卷积神经网络通过加入时间维度来代替光流，可以实现端到端的识别。Tran等[14]使用3D卷积构建了C3D（Convolutional Three Dimensional）模型，它将VGGNet（Visual Geometry Group network）网络[15]的卷积核由3×3的2D卷积扩展为3×3×3的3D卷积。之后出现的R3D（Residual 3D Convolutional Network）[16]网络和SlowFast[17]网络等都基于3D卷积神经网络。此外，在长短时记忆网络（Long Short-Term Memory, LSTM）的进展中，Donahue等[18]引入了长期循环卷积神经网络（Long-term recurrent Convolutional Networks, LRCN）的概念。LRCN结合了2D卷积神经网络（2D CNN）来提取帧级特征，并随后利用LSTM来建模多个视频帧之间的时间关系。这一方法在视频行为识别领域具有重要的应用潜力。

上述研究方法在区分设施黄瓜的生长过程中的复杂农事行为时，面临着一系列挑战，包括株距较近、叶片相互遮挡、农事操作多样、动作环节复杂以及人员操作不规范等问题。这些问题增加了设施黄瓜的农事行为识别的难度。为了解决这些挑战，本研究基于SlowFast行为识别算法进行了改进。具体地，在Fast Pathway中将ACTION（Spatio-temporal, Channel and Motion Excitation）[19]注意力机制与残差块相结合，形成SMC-Res Block，以增强相邻两帧之间农事操作的连续性特征提取能力。考虑黄瓜生产中叶片遮挡和大棚环境的复杂性，在Slow Pathway中引入了注意力机制ECANet（Efficient Channel Attention Network），以增强通道之间的相互依赖关系，从而提高Slow Pathway网络的特征表示能力。此外，为解决农事行为数据集中的不均衡问题，本研究设计了平衡损失函数（Smoothing Loss, SLoss）。使用这一损失函数有助于平衡各个农事行为类别在数据集中的样本分布，从而提高模型对于每个类别的识别性能。

1　农事行为数据集构建

鉴于当前缺乏适用于种植黄瓜的农事行为监控的公开可用的数据集，本研究选用北京国家精准农业实验示范基地内的黄瓜温室为研究案例，并自行构建数据集，用于识别和评价种植黄瓜的农事行为。为了确保能够捕捉到农业操作人员的动作，研究采用以下布置方式：温室的长宽比为A∶B

(A>B，其中A为长15 m，B为宽3 m)，行距为100 cm，株距为40 cm，共有18垄，垄间距为80 cm。根据这一布局，摄像头的安装点位如图1所示，摄像头被设置在长边上，每两垄黄瓜苗之间，以确保清晰拍摄农业操作人员的操作。考虑监控视频的主要目的是识别农业人员与黄瓜的互动行为，摄像头的安装高度为2.2 m，略高于人的头顶高度。此外，摄像头角度倾斜15～30º，以确保能够清晰捕捉操作人员的行为。为增加角度的多样性，还使用手机对农事行为进行辅助拍摄。

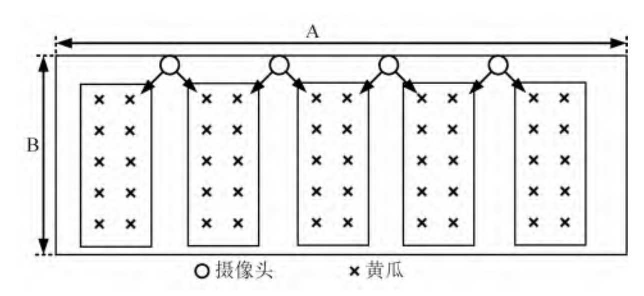


图1 农事行为识别研究摄像头布置点位

Fig. 1 The arrangement of camera positions for agricultural activity recognition research

实验数据的采集时间为2023年2月28日—4月25日。此过程为黄瓜从移栽到采摘的全部过程。拍摄的设备为海康威视的萤石云家用摄像头，型号为DS-IPC-B12V2-I/H8，焦距为6 mm，清晰度为5 MP，拍摄的时间为10：00～11：30和14：00～15：00，每段视频拍摄的时长大约为1 min。本研究共采集了移栽、喷水、吊蔓、整枝、采摘共5个行为197段视频。图2所示为采集到的部分农事行为的视频截帧。拍摄视频的标识方法为AVA数据集格式[16]。视频数据经过切分与筛选，农事行为数据集一共有707段视频，其中500段视频用作训练集；40段视频用作验证集；67段视频用作测试集。

在建立原始数据集后，对数据进行抽帧和标注。为了确保数据的均衡性，每个行为的数据量需相当，并且不能截断任何动作。为实现这一目标，采取了以下措施：1）删除视频中没有目标人员出现的片段。2）将视频中包含目标前后多个动作的片段进行拆分。最终的数据集组成如表1所示。

2 模型构建

本研究提出的SlowFast-SMC-ECA模型基于SlowFast模型。其结构如图3所示，主要包括数据层、卷积层、残差层及特征融合层。模型的整体处理流程：数据层通过2个不同的步长值得到不同帧



a.采摘行为



b.吊蔓行为

图2 设施黄瓜的农事行为视频截帧

Fig. 2 Video cut-off frames of agronomic behaviour of facility cucumbers

表1 设施黄瓜农事行为数据集的构成

Table 1 The composition of the greenhouse cucumber farming behavior dataset

行为类别	视频数/个	标签数量/个
移栽	97	7 432
浇水	146	10 207
吊蔓	162	11 106
整枝	124	9 978
采摘	178	12 173

的数据将其馈送到不同的通道中，在进入到卷积层后，Slow Pathway每次以1帧进行运算；Fast Pathway提取5帧图片一起进行运算。接着进入3D残差网络，Slow Pathway和Fast Pathway分别用ECA-Res和多路径激励残差网络进行农事活动行为中运动信息和空间信息的提取，最后进行特征融合，得到最终的农事活动行为的结果。

2.1 多路径激励残差网络

在Fast Pathway中以高时间分辨率捕获运动信息，但是基于设施黄瓜的农事活动行复杂多变，手部动作幅度小，一些农事行为相关性强，原始残差块在捕获农事活动行为运动特征时会丢失大量信息，造成误检现象。本研究利用ACTION[19]中的3个互补注意力机制，即STE（Spatial-Temporal Excitation）、CE（Channel Excitation）、ME（Motion Excitation），结合原始的残差块，形成多路径激励残差网络（Spatial-Temporal Excitation、Channel

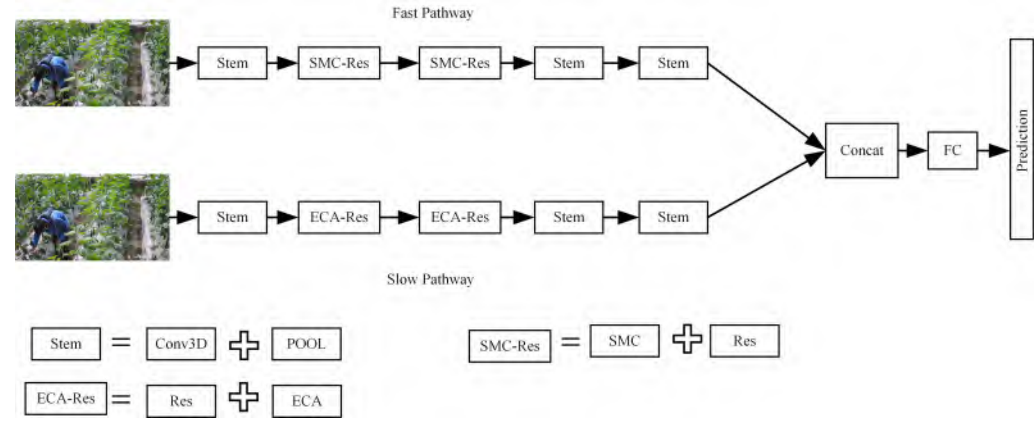


图3　SlowFast-SMC-ECA网络结构图

Fig. 3　SlowFast-SMC-ECA network structure diagram

Excitation、Motion Excitation Residual，SMC-Res，结构如图4所示）来提高对农事行为视频中关键特征的激发，从而提升农事行为识别的准确性。本研究使用Conv为卷积数量；F为内核大小；卷积滤波器的特征映射数为 $n \times x \times x$ 和 F；BN（Batch Normalization）为批量归一化；ReLu为激活函数。SMC-Res块包含3次卷积和一个残差边，在每一组卷积以及残差连接之前加入SMC模块。这样做可以在不同维度获取多类型的时空模式、通道信息及运动信息后进行卷积获取更加细粒度的特征，提高农事行为识别的精度。

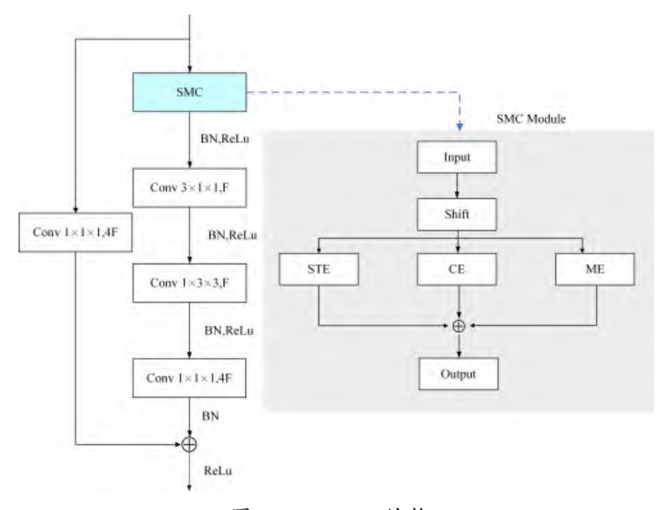


图4　SMC-Res结构

Fig. 4　SMC-Res structure

STE时空注意力模块旨在捕获适当的时空模式，以强化视频中的空间和时间关系，有助于更好地理解农事活动行为中的一些动作的变化，主要通过生成时空掩码来产生时空注意力图，以提取视频中的时空特征。STE网络结构如图5a所示。

首先对输入 $\boldsymbol{X} \in \boldsymbol{R}^{N \times T \times C \times H \times W}$（$N$表示批量大小；$T$表示段数；$C$表示通道数；$H$表示高度；$W$表示宽度）做一个通道平均得到关于通道轴的全局时空张量 $\boldsymbol{F}_1 \in \boldsymbol{R}^{N \times T \times 1 \times H \times W}$，后把得到的 $\boldsymbol{F}_1$ 重构为新的时空张量 $\boldsymbol{F}_1^* \in \boldsymbol{R}^{N \times 1 \times T \times H \times W}$，然后馈送到3D卷积核$K$中，数学表达如公式（1）所示。

$$\boldsymbol{F}_{o1}^* = K \times \boldsymbol{F}_1^* \tag{1}$$

然后再将 $\boldsymbol{F}_{o1}^*$ 重构为 $\boldsymbol{F}_{o1}$，最后经过Sigmoid函数进行激活得到掩码，如公式（2）所示。

$$\boldsymbol{M}_1 = \delta(\boldsymbol{F}_{o1}) \tag{2}$$

式中：M_1为激活掩码；δ为Sigmoid函数。最后得到农事活动行为中更为精细的时空信息，如公式（3）所示。

$$Y_1 = X + X \odot M_1 \tag{3}$$

式中：Y_1为STE模块的最终输出。

CE注意力模块用于提取适当的通道范围特征，以强调网络中不同通道的信息。这有助于捕捉关键的通道信息，从而提高农事行为的识别能力。它类似于SE（Squeeze-and-Excitation Networks）注意力模块[20]，为了增强农事活动行为各个不同特征在时间上的相互依赖程度，CE模块在两个全连接层之间引入了一个一维卷积层，以捕捉信道特征上的时间信息。CE模块的结构如图5b所示。对于给定的输入 $\boldsymbol{X} \in \boldsymbol{R}^{N \times T \times C \times H \times W}$ 通过平均池化的方法来获取全局空间信息 $\boldsymbol{F}_2 \in \boldsymbol{R}^{N \times T \times C \times 1 \times 1}$，它的求解如公式（4）所示。

$$\boldsymbol{F}_2 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{X}[:,:,:,i,j] \tag{4}$$

对 $\boldsymbol{F}_2$ 用一个二维卷积核 K_1 进行压缩得到 $\boldsymbol{F}_r$，如公式（5）所示。

$$\boldsymbol{F}_r = K_1 \times \boldsymbol{F}_2 \tag{5}$$

对 $\boldsymbol{F}_r \in \boldsymbol{R}^{N \times T \times \frac{C}{r} \times 1 \times 1}$ 重构得到 $\boldsymbol{F}_r^* \in \boldsymbol{R}^{N \times \frac{C}{r} \times T \times 1 \times 1}$，

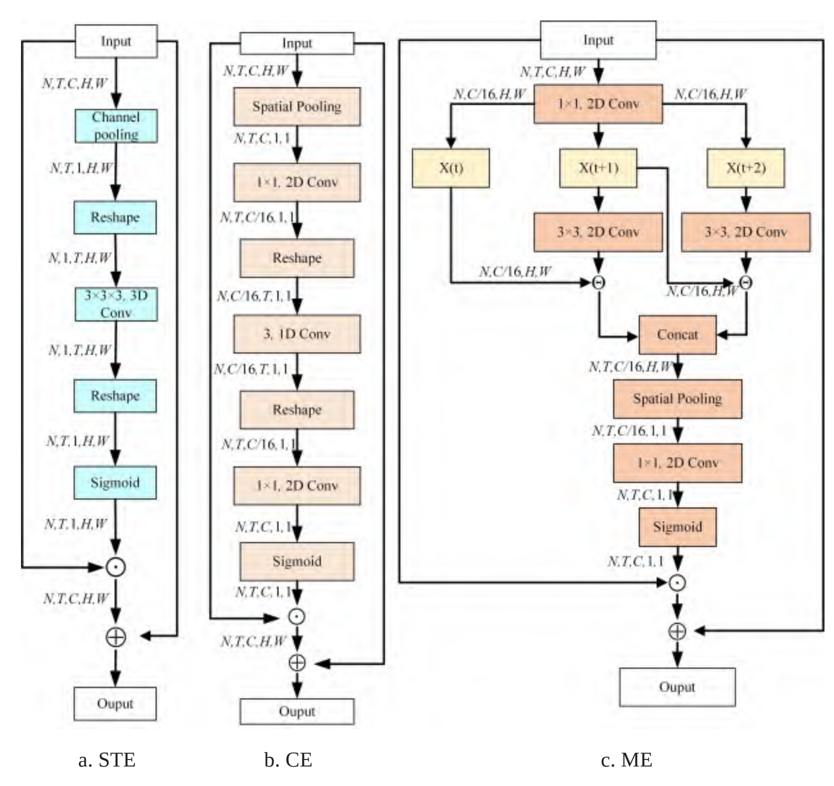


图5 SMC模块组成

Fig. 5 Composition of SMC modules

然后用核大小为3的一维卷积核K_2与F_r^*相乘得到$F_{temp}^* \in R^{N \times \frac{C}{r} \times T \times 1 \times 1}$，然后重构$F_{temp}^*$得到$F_{temp}$后与2D卷积核$K_3$相乘进行解压缩，最后经过Sigmoid函数得到农事动作的掩码，如公式（6）和公式（7）所示。

$$M_2 = \delta(F_{o2}) \quad (6)$$

$$Y_2 = X + X \odot M_2 \quad (7)$$

式中：M_2为CE模块的掩码；Y_2为最终输出。

ME注意力模块专注于提取运动信息，以更好地聚焦于农事活动行为中操作人员的手部动作的变化，如图5c表示，通过相邻帧之间的变化情况来建模农事活动行为的运动特征，如公式（8）所示。

$$F_m = K \times F_r[:, t+1, :, :, :] - F_r[:, t, :, :, :] \quad (8)$$

式中：K为3×3的二维卷积；F_m通过K对前后两帧的操作得到，即将输入X每相邻两帧之间得到的差值在时间维度上进行连接。再对得到的特征做平均池化处理，然后通过Sigmoid函数得到最终的输出，如公式（9）所示。

$$Y_3 = X + X \odot M_3 \quad (9)$$

式中：M_3为ME模块的掩码；Y_3为最终输出。

农事行为的特征信息通过STE、CE、ME注意力机制，将生成的3个激发特征逐元素相加，再经过多路径激励通道，最终结果如公式（10）所示。

$$Y = Y_1 + Y_2 + Y_3 \quad (10)$$

式中：Y为多路径激励残差网络的最终输出。

然后对生成的特征信息进行卷积操作，得到最终的运动特征信息。

2.2 ECA-Res残差块

在设施黄瓜的生产环境下，黄瓜在开花期叶片生长迅速，存在黄瓜叶片遮挡操作人员的手部动作变化的问题。在Slow Pathway中，如操作人员的手部轮廓等通道信息不容易被捕捉到。为提高通道信息的捕捉能力，在ResNet主干网络的基础上，在残差块中结合ECA注意力[21]。ECANet注意力机制在SENet注意力机制的基础上实现了不降维的跨通道交互策略，只涉及了少量的参数，不仅避免了维度特征的缩减，还能增加通道之间的信息交互，在保证交互的前提下精简模型。SlowFast网络中的Slow Pathway有比Fast Pathway更多的通道数量来学习通道信息。ECA注意力机制可以完美地适用于SlowFast网络中的Slow Pathway，不仅减少了计算量，还突出了通道中的关键信息和抑制视频中背景因素的干扰，其结构如图6所示。

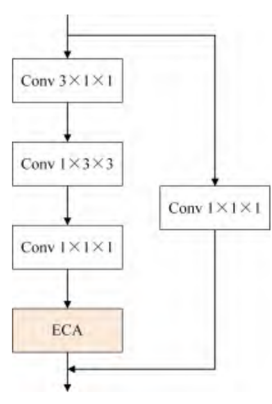


图6 ECA-Res结构

Fig. 6 ECA-Res structure

ECA注意力机制的工作原理如图7所示。通过卷积对特征图进行压缩得到一个新的特征图χ。χ的大小为$H \times C \times W$。将经过全局平均池化（Golbal Average Pooling, GAP）转变为$1 \times 1 \times C$的向量。这样空间信息就得到了压缩，然后采用一维卷积来提取通道上的特征，模型在训练的过程中，能够自适应卷积核的大小，具体的做法为：

1）在全局平局池化之后得到一个1 × 1 × C的向量。

2）计算自适应一维卷积核的大小如公式（11）所示。

$$k = \varphi(c) = \left| \frac{\log_2(c)}{\gamma} + \frac{b}{\gamma} \right| \quad (11)$$

式中：$\gamma = 2$； $b = 1$；k为核大小；c为通道大小。该自适应卷积核表明了局部跨通道交互的覆盖率。

3）将自适应卷积核使用到一维卷积中，得到各通道的权重，使得通道数较大的层可以更多地进行相邻通道间的交互。

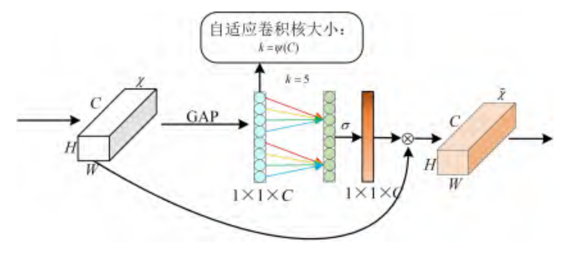


图7　ECANet网络结构图

Fig. 7　ECANet network structure diagram

2.3 损失函数

在数据采集过程中，由于黄瓜生长周期中不同农事行为的频率差异，某些行为（如移栽）在整个生长周期中仅发生几次，而其他行为（如浇水、采摘）则有较高的发生频率，这导致数据集存在明显的类别不均衡问题。同时一些行为（如吊蔓和采摘）在表现上相似。所以原始的损失函数[22]在对于一些小样本的农事行为活动时，它的准确率得不到保证。为解决这一问题，本研究设计了一种平衡损失函数（Smoothing Loss, SLoss）。该损失函数通过引入正则化系数α^i并与原始损失函数相乘，旨在缓解模型在训练过程中对于高频行为的过拟合，同时确保对于低频行为的充分训练。正则化系数的定义如公式（12）所示。

$$\alpha^i = S(1 - \chi^i) \quad (12)$$

式中：S（一般设置为0.1、0.5、0.75。本研究设置为0.75）用来平衡类别的超参数；χ^i的计算如公式（13）所示。

$$\chi^i = \frac{f_i}{\sum_{i=1}^{C} f_i} \quad (13)$$

式中：f_i为第i类农事行为的样本个数；C为农事行为类别数。因此，最后的SLoss如公式（14）所示。

$$\text{SLoss} = \alpha^i \text{Loss}(p_t) = S(1 - \chi^i)\text{Loss}(p_t) \quad (14)$$

式中：p_t为预测为正类别的概率。

在多分类任务中，经过Sigmoid函数进行归一化处理后得到最终的结果，如公式（15）所示。

$$\begin{aligned} \text{SLoss}(p_t) &= -\sum_{i=1}^{C} \alpha_i \log[\text{sigmoid}(p_t)] \\ &= -\sum_{i=1}^{C} S(1-\chi^i)\log\left[\frac{1}{1+\exp(-p_t)}\right] \\ &= -\sum_{i=1}^{C} S(1-\chi^i)\log(p_t) \end{aligned} \quad (15)$$

3 实验设计与结果分析

3.1 实验环境

本研究的实验环境为Linux 5的操作系统，CPU为Intel（R）Xeon（R）Platinum 8255C CPU@ 2.50 GHz，GPU为NVIDIA GeForce GTX 2080 Ti显卡，深度学习的框架为PyTorch框架。

在模型的训练过程中，模型训练的Epoch设置为200，批量大小设置为8，初始的学习率为0.001，权重衰减参数设置为0.005。网络中的模型优化采用的是随机梯度下降算法（Stochastic Gradient Descent, SGD）。

3.2 实验结果与分析

3.2.1 农事活动行为识别结果

为验证本研究提出的改进SlowFast模型的农事行为识别方法对移栽、浇水、吊蔓、整枝、采摘5种行为识别效果的优越性，将SlowFast-SMC-ECA模型与原模型SlowFast进行比较。对比结果如表2所示。

表2　SlowFast-SMC-ECA模型不同行为识别精度对比

Table 2　Comparison of accuracy in different behavior recognition of SlowFast-SMC-ECA model

行为类别	mAP@0.5/%	
	SlowFast-SMC-ECA	SlowFast
移栽	76.85	73.32
浇水	82.28	80.76
吊蔓	78.23	75.32
整枝	77.65	74.43
采摘	86.61	85.96
全部行为	80.47	77.87

由表2可以看出，改进后方法相比原始的SlowFast模型在5种农事行为的识别准确率均有不同幅度的提升。其中，提升较为明显的是移栽行为，较原模型提高3.53%。提升不太明显的是采摘行为，仅为86.61%。全部行为识别精度的平均值为

80.47%，较原模型提高2.6%。

在本研究的方法中，对于吊蔓和整枝这两种行为的识别能力相对较低，识别精度仅有78.23%和77.65%。分析其原因是这两种行为表现比较相似，模型容易产生混淆，但是相对于原模型的75.32%和74.43%，识别精度都提高大约3%。可见本研究的方法对于吊蔓和整枝这两种易混淆行为的识别能力也有显著的增强。对于移栽这种样本量比较小的行为，它的mAP值能够提升大约3.5%。可见模型对于小样本数目的类别有一个较好的优化。图8为SlowFast-SMC-ECA模型农事行为识别的结果图。



a. 移栽行为 b. 浇水行为 c. 吊蔓行为 d. 整枝行为 e. 采摘行为

图8 不同黄瓜农事行为检测视频帧结果图

Fig. 8 Different cucumber farming activity detection frame results

3.2.2 消融实验

为了验证在不同阶段使用SMC-Res和ECA-Res残差网络的效果差异，对Res1、Res2、Res3、Res4和Res5这5个阶段进行了实验。在每个阶段，分别用SMC-Res和ECA-Res替换原来的残差网络，并在第1阶段完成后，不将其恢复为原始的残差块，而是直接基于此结果将第2阶段的原始残差块替换为多路径激励残差网络和ECA-Res。随后的阶段也以同样的操作方式将原始的残差块替换为i多路径激励残差网络和ECA-Res，结果如表3所示。

实验结果表明，将Res2和Res3这两个残差网络替换为SMC-Res和ECA-Res的效果较好。这是由于在Res2和Res3中分别包含3个SMC-Res残差块和4个ECA-Res。它的网络输出的特征图有更多的信息和强大的空间相关性，在这里进行操作可以有效地防止过拟合，同时网络可以更好地提取空间信息。在Res1、Res4、Res5之后添加几乎没有效果。前者是因为在经过一层卷积过后视野太大，提取的特征不够充分，将原始残差网络替换并不能有效地提取农事行为的特征信息；后者是因为深层的卷积神经网络输出的特征图的相关性较弱，再次执行SMC-Res和ECA-Res操作后，会丢失过多的农事行为特征信息，不利于网络更好地学习。因此，本研究只将Res2和Res3替换为SMC-Res和ECA-Res残差网络。

表3 SlowFast-SMC-ECA模型不同位置残差块的实验效果图

Table 3 Experimental results of different position residual blocks in SlowFast-SMC-ECA model

不同位置的残差块	mAP@0.5/%
SMC-Res1+ECA-Res1	76.93
SMC-Res2+ECA-Res2	78.55
SMC-Res3+ECA-Res3	80.03
SMC-Res4+ECA-Res4	79.97
SMC-Res5+ECA-Res5	77.75

同时，为验证本研究提出的农事行为活动识别方法对原模型改进的有效性，对SlowFast、SlowFast+SMC、SlowFast+ECA、SlowFast+SLoss、SlowFast+SMC+ECA这5个模型，通过消融实验对识别效果进行对比，进一步验证本研究模型的实验效果的性能，结果如表4所示。

表4 农事行为识别研究消融实验效果表

Table 4 Dissolution experiment results of agricultural activity recognition research

模型	mAP@0.5/%
SlowFast	77.87
SlowFast+SMC-Res	78.55
SlowFast+ECA-Res	78.32
SlowFast+SLoss	78.25
SlowFast+SMC-Res+ECA-Res	80.18
SlowFast-SMC-ECA	80.47

根据表4的结果，通过将SlowFast模型中的原始Res残差块替换为SMC-Res和ECA-Res残差块，明显提升了农事行为识别效果，达到80.18%，相较于原模型SlowFast的识别精度提高了约2%。值得注意的是，平衡损失函数对整体农事行为识别效果的提升并不十分显著，仅为0.38%。然而，在处理

小样本的行为时，平衡损失函数却表现出较大的提升效果。综合而言，SMC-Res、ECA-Res的引入及对损失函数的改进，有效提升了农事行为识别的准确性。

3.2.3　对比实验

为了验证本研究的网络模型性能，将本研究的模型与其他行为识别模型在农事行为数据集上进行实验。本次实验的网络模型主要有C3D、I3D、TSN、双流卷积神经网络、Timesformer（Time-space Transformer），以及本研究的网络模型。各个模型的平均识别准确率如表5所示。

表5　农事行为识别研究的对比实验效果表

Table 5 Comparative experimental results table of agricultural activity recognition research

模型	mAP@0.5/%
C3D	78.78
I3D	77.89
TSN	78.56
双流卷积神经网络	75.45
Timesformer	79.47
SlowFast-SMC-ECA	80.47

图9是原始的SlowFast和SlowFast-SMC-ECA这2种模型的训练损失率变化曲线。图9中横坐标为迭代的次数，纵坐标为损失率，可以看出SlowFast-SMC-ECA模型经过120次迭代后收敛到了0.03且模型基本收敛，而原始的SlowFast模型需要经过160次的迭代才能基本收敛。从表5可以看出，本研究方法与其他几个方法进行比较，平均识别的准确率最高。由此可知，本研究的方法在识别效果上优于其他方法，并且改进后的模型收敛的速度更快，效果更好。

4　结论与讨论

为了能够准确快速地识别农事行为活动，本研究提出了一种改进的SlowFast农事活动行为识别算法，主要结论如下。

1）自建了一个关于设施黄瓜的农事活动行为数据集，包括移栽、浇水、吊蔓、整枝和采摘这5种农事活动行为。

2）为解决农事活动行为动作复杂且设施环境复杂的问题，本研究在原模型的基础上进行了改进，具体做法包括在Fast Pathway中结合ACTION注意力机制和残差块，形成了SMC-RES残差网络，以增强对农事操作信息的提取；在Slow Pathway中

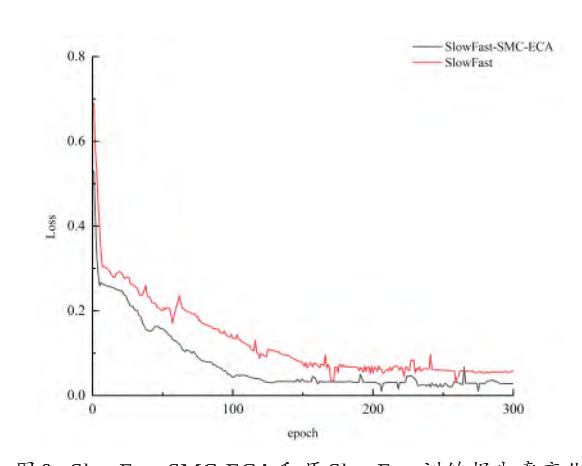


图9　SlowFast-SMC-ECA和原SlowFast训练损失率变化

Fig. 9 SlowFast-SMC-ECA and original SlowFast training loss rate changes

引入ECA结构，提高了农业人员位置、大小等空间语义信息的提取，以此来提高农事行为识别的准确性。

3）为解决数据集中农事行为类别不平衡的问题，本研究设计了平衡损失函数（Smoothing Loss），用于保证对于低频农事行为的充分训练及防止对于高频农事行为的过拟合。

经过实验，SlowFast-SMC-ECA模型相较于原始网络模型提高约2%的mAP，实验证实了SMC-Res残差网络、ECA-Res残差块和平衡损失函数对SlowFast模型的改进效果和识别准确性的提升。尽管在改进过程中仍存在误检现象，同时由于SlowFast模型参数较多，难以嵌入监控设备中，但这一研究在一定程度上推动了农事行为识别的进一步研究。未来的工作将继续改进模型，使其更加准确和轻量化，以便在记录农业人员从事农业活动的同时，有效记录农事行为。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 武春成, 周国彦, 曹霞, 等. 连作土壤连续施入生物炭对黄瓜品质及根区微生态的影响[J]. 江苏农业科学, 2022, 50(9): 143-147.
WU C C, ZHOU G Y, CAO X, et al. Influences of continuous application of biochar in continuous cultivated soils on cucumber quality and root zone micro ecology[J]. Jiangsu agricultural sciences, 2022, 50(9): 143-147.

[2] 尚小红, 周生茂, 郭元元, 等. 黄瓜异根嫁接植株抗逆性变化研究进展[J]. 中国细胞生物学学报, 2017, 39(3): 364-372.
SHANG X H, ZHOU S M, GUO Y Y, et al. Advances in stress-resistant changes in hetero-grafting cucumber (*Cucumis sativus* L.) plant[J]. Chinese journal of cell biology,

2017, 39(3): 364-372.

[3] 童安炀, 唐超, 王文剑. 基于双流网络与支持向量机融合的人体行为识别[J]. 模式识别与人工智能, 2021, 34(9): 863-870.
TONG A Y, TANG C, WANG W J. Human action recognition fusing two-stream networks and SVM[J]. Pattern recognition and artificial intelligence, 2021, 34(9): 863-870.

[4] WANG C, WANG S F, LI J J, et al. Research on the identification method of overhead transmission line breeze vibration broken strands based on VMD-SSA-SVM[J]. Electronics, 2022, 11(19): ID 3028.

[5] ZHANG X K, WANG Y J, DOU Z H, et al. Residual current fault type recognition based on S3VM and KNN cooperative training[J]. Journal of power electronics, 2022, 22(11): 1966-1977.

[6] 张春, 郭明亮. 大数据环境下朴素贝叶斯分类算法的改进与实现[J]. 北京交通大学学报, 2015, 39(2): 35-41.
ZHANG C, GUO M L. Research and realization of improved native Bayes classification algorithm under big data environment[J]. Journal of Beijing jiaotong university, 2015, 39(2): 35-41.

[7] LETHIKIM N, NGUYENTRANG T, VOVAN T. A new image classification method using interval texture feature and improved Bayesian classifier[J]. Multimedia tools and applications, 2022, 81(25): 36473-36488.

[8] PATIÑO-SAUCEDO J A, ARIZA-COLPAS P P, BUTT-AZIZ S, et al. Predictive model for human activity recognition based on machine learning and feature selection techniques[J]. International journal of environmental research and public health, 2022, 19(19): ID 12272.

[9] WANG H, KLÄSER A, SCHMID C, et al. Action recognition by dense trajectories[C]// CVPR. Piscataway, New Jersey, USA: IEEE, 2011: 3169-3176.

[10] WANG H, SCHMID C. Action recognition with improved trajectories[C]// 2013 IEEE International Conference on Computer Vision. Piscataway, New Jersey, USA: IEEE, 2013: 3551-3558.

[11] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 1. New York, USA: ACM, 2014: 568-576.

[12] WANG L M, XIONG Y J, WANG Z, et al. Temporal segment networks: Towards good practices for deep action recognition[C]// European Conference on Computer Vision. Berlin, German: Springer, 2016: 20-36.

[13] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? A new model and the kinetics dataset[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2017: 4724-4733.

[14] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks[C]// 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2015: 4489-4497.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. arXiv: 1409.1556, 2014.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 770-778.

[17] FEICHTENHOFER C, FAN H Q, MALIK J, et al. SlowFast networks for video recognition[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2019: 6201-6210.

[18] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2015: 2625-2634.

[19] WANG Z W, SHE Q, SMOLIC A. ACTION-net: Multipath excitation for action recognition[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13209-13218.

[20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 7132-7141.

[21] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 11531-11539.

[22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2999-3007.

Recognition Method of Facility Cucumber Farming Behaviours Based on Improved SlowFast Model

HE Feng[1,2], WU Huarui[1,2,3,4], SHI Yangming[1,2], ZHU Huaji[1,2,3,4*]

(1. *School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang* 212013, *China*; 2. *National Engineering Technology Research Centre for Agricultural Informatisation, Beijing* 100097, *China*; 3. *Information Technology Research Centre, Beijing Academy of Agriculture and Forestry, Beijing* 100097, *China*; 4. *Key Laboratory of Digital Rural Technology, Ministry of Agriculture and Rural Development, Beijing* 100097, *China*)

Abstract:

[Objective] The identification of agricultural activities plays a crucial role for greenhouse vegetables production, particularly in the precise management of cucumber cultivation. By monitoring and analyzing the timing and procedures of agricultural operations, effective guidance can be provided for agricultural production, leading to increased crop yield and quality. However, in practical applications, the recognition of agricultural activities in cucumber cultivation faces significant challenges. The complex and ever-changing growing environment of cucumbers, including dense foliage and internal facility structures that may obstruct visibility, poses difficulties in recognizing agricultural activities. Additionally, agricultural tasks involve various stages such as planting, irrigation, fertilization, and pruning, each with specific operational intricacies and skill requirements. This requires the recognition system to accurately capture the characteristics of various complex movements to ensure the accuracy and reliability of the entire recognition process. To address the complex challenges, an innovative algorithm: SlowFast-SMC-ECA (SlowFast-Spatio-Temporal Excitation, Channel Excitation, Motion Excitation-Efficient Channel Attention) was proposed for the recognition of agricultural activity behaviors in cucumber cultivation within facilities.

[Methods] This algorithm represents a significant enhancement to the traditional SlowFast model, with the goal of more accurately capturing hand motion features and crucial dynamic information in agricultural activities. The fundamental concept of the SlowFast model involved processing video streams through two distinct pathways: the Slow Pathway concentrated on capturing spatial detail information, while the Fast Pathway emphasized capturing temporal changes in rapid movements. To further improve information exchange between the Slow and Fast pathways, lateral connections were incorporated at each stage. Building upon this foundation, the study introduced innovative enhancements to both pathways, improving the overall performance of the model. In the Fast Pathway, a multi-path residual network (SMC) concept was introduced, incorporating convolutional layers between different channels to strengthen temporal interconnectivity. This design enabled the algorithm to sensitively detect subtle temporal variations in rapid movements, thereby enhancing the recognition capability for swift agricultural actions. Meanwhile, in the Slow Pathway, the traditional residual block was replaced with the ECA-Res structure, integrating an effective channel attention mechanism (ECA) to improve the model's capacity to capture channel information. The adaptive adjustment of channel weights by the ECA-Res structure enriched feature expression and differentiation, enhancing the model's understanding and grasp of key spatial information in agricultural activities. Furthermore, to address the challenge of class imbalance in practical scenarios, a balanced loss function (Smoothing Loss) was developed. By introducing regularization coefficients, this loss function could automatically adjust the weights of different categories during training, effectively mitigating the impact of class imbalance and ensuring improved recognition performance across all categories.

[Results and Discussions] The experimental results significantly demonstrated the outstanding performance of the improved SlowFast-SMC-ECA model on a specially constructed agricultural activity dataset. Specifically, the model achieved an average recognition accuracy of 80.47%, representing an improvement of approximately 3.5% compared to the original SlowFast model. This achievement highlighted the effectiveness of the proposed improvements. Further ablation studies revealed that replacing traditional residual blocks with the multi-path residual network (SMC) and ECA-Res structures in the second and third stages of the SlowFast model leads to superior results. This highlighted that the improvements made to the Fast Pathway and Slow Pathway played a crucial role in enhancing the model's ability to capture details of agricultural activities. Additional ablation studies also confirmed the significant impact of these two improvements on improving the accuracy of agricultural activity recognition. Compared to existing algorithms, the improved SlowFast-SMC-ECA model exhibited a clear advantage in prediction accuracy. This not only validated the potential application of the proposed model in agricultural activity recognition but also provided strong technical support for the advancement of precision agriculture technology. In conclusion, through careful refinement and optimization of the SlowFast model, it was successfully enhanced the model's recognition capabilities in complex agricultural scenarios, contributing valuable technological advancements to precision management in greenhouse cucumber cultivation.

[Conclusions] By introducing advanced recognition technologies and intelligent algorithms, this study enhances the accuracy and efficiency of monitoring agricultural activities, assists farmers and agricultural experts in managing and guiding the operational processes within planting facilities more efficiently. Moreover, the research outcomes are of immense value in improving the traceability system for agricultural product quality and safety, ensuring the reliability and transparency of agricultural product quality.

Key words: farming activity behaviour; SlowFast model; multi-path incentive residual network; ECA-Res; equilibrium loss function

Foundation items: Central Guided Local Science and Technology Development Funds Project (2023ZY1-CGZY-01); Ministry of Finance and Ministry of Agriculture and Rural Development: Funding for the National Modern Agricultural Industrial Technology System (CARS-23-D07)

Biography: HE Feng,1363263324@qq.com

***Corresponding author:** ZHU Huaji, zhuhj@nercita.org.cn